Introduction to EAC-CPF 2.0

Transcription of 2021-04-27 webinar hosted by the Society of American Archivists' (SAA) Technical Subcommittee on Encoded Archival Standards (TS-EAS), in collaboration with the Encoded Archival Standards Section Steering Committee

---

first slide

[Speaker: Cory Nimer] We will be recording the presentation portion of today's session but not the Q&A, and posting it to SAA's YouTube channel.

Today's session is meant to introduce the changes that have been proposed in the draft version of EAC-CPF. I'd like to introduce our presenters:

- Silke Jagodzinski
  EAC-CPF Subteam, Chair
  Prussian Secret State Archives

- Kerstin Arnold
  EAD Subteam, Chair
  Archives Portal Europe Foundation

- Ailie Smith
  EAC-CPF Subteam member
  The University of Melbourne

- Karin Bredenberg
  TS-EAS, Co-chair
  Kommunalforbundent Sydarkivera

At this point we'll turn the presentation over to our presenters.

[Speaker: Karin Bredenberg] Thank you, Cory. I will start off with some minor remarks regarding TS-EAS.

next slide

As you heard, the Technical Subcommittee on Encoded Archival Standards is at the Society of American Archivists. And luckily for you, we have a presentation and webinar already recorded (https://youtu.be/9NXNyx9py-I), giving you all of the background information you need. So I'm just going to say: we take care of all of the formats that you use for managing and sharing archival information, and our work is driven by comments, suggestions, and bug

reports. That's how we evolve the standards; it's not something we're sitting and doing on our own. It's your input that drives our work forward.

next slide

And of course, we are available in a lot of places. We have one page for the whole committee (https://www2.archivists.org/groups/technical-subcommittee-on-encoded-archival-standards-ts-eas). We are maintaining the standards on GitHub (https://github.com/SAA-SDT) -- so there is one way to access them. But we also have publications both of the EAD tag library and schema, and the EAC tag library and schema (http://www.loc.gov/ead/index.html and https://eac.staatsbibliothek-berlin.de/). We do have one mailing list for all of the standards that we are working with: the EAD mailing list at the Library of Congress (https://www.loc.gov/ead/eadlist.html). And we also have -- besides using GitHub -- a form you can use to report an issue (https://www2.archivists.org/standards/TS-EAS-report-an-issue).

But the main thing today is why we are having this session.

next slide

We are talking about the standards revision. We do minor changes during a normal year--we have a set revision cycle and manifest for it, available in GitHub. But we haven't just created the standards to do minor updates. Following the guidelines that the Standards Committee hosted by SAA--of which we are a subcommittee, too--every standard needs to go through a revision every fifth year: a five-year cycle for minor changes, then the big revision starts to look over the standard standard to see if major changes are needed.

With that, that's the work that the EAC-CPF has been doing for the last couple of years. It takes time, we gather input--and we are currently gathering input. To familiarize you with this revision, I'm going to hand it over to the EAC-CPF team, and let them take us through this revision. Thank you very much to the EAC-CPF team, and the floor is yours!

next slide

[Speaker: Silke Jagodzinski] Thanks, Karin. I'd like to begin by introducing my co-presenters:

next slide

Ailie Smith is Digital Curation and Archives Specialist in the Digital Stewardship team at the University of Melbourne in Australia. Ailie has attended countless virtual team meetings monthly on her Friday nights. She has worked to improve the Tag Library and the Documentation and she is an expert in encoding dates in XML.

[Speaker: Ailie Smith] Thank you, Silke. I'd like to introduce Silke Jagodzinski, who is Head of the Digital Services and IT Unit at the Geheimes Staatsarchiv Preußischer Kulturbesitz, in Berlin, Germany. Silke is the team lead for the EAC-CPF team and has been a driving force in getting this revision of EAC-CPF to where it is today.

next slide

[Speaker: Silke Jagodzinski] Our latest and current version of the EAC-CPF XML schema was approved in 2010. In summer 2017 at the SAA Annual meeting in Portland, Oregon, the TS-EAS decided to revise the standard version EAC-CPF 2010. A call for comments followed in September 2017, and the community replied with feedback on bugs and requests for simplifying the encoding and for enhancement of the description, especially the Tag Library.
A technical update was published at the end of 2018, which is why the standard is called EAC-CPF 2010 edition 2018. After the 2018 update, a major overhaul of the standard started in the EAC-CPF team, consisting of 5 to 10 members, with the target of a revised EAC-CPF 2.0 schema. We will present parts of the revised draft during the next 20 minutes.

[The EAC-CPF team has monthly virtual meetings and, until 2019, yearly in person meetings during SAA Annual meeting and we had an additional Revision meeting in March 2020 in Berlin.]

next slide

During our meetings, each element and each attribute of the schema was evaluated to

- simplify where it is possible,
- to align with EAD schema where it is useful,
- to implement features and solutions upon our users' request, and
- to clear up unused components.

Since March this year the EAC-CPF 2.0 draft has been published with a Call for Comments.

next slide

In the Call for Comments, we provide schema files (xsd, rng), schematron, a conversion stylesheet, a Tag Library without examples as well as revision notes and a questionnaire with a list of topics we are deeply interested in to find use cases or published examples.

next slide

The call for comments closes at the end of June 2021. We are going to discuss the feedback and proposals and will include the results in the schema and documentation. We are going to produce a Best Practices Guide and aim to release the new EAC-CPF 2.0 version at the end of 2021 and you can find all this information on our EAC-CPF homepage.

And now I'd like to hand over to Ailie for more about simplifying the schema and EAD reconciliation.

[Speaker: Ailie Smith] I'll start by talking briefly about the design principles (https://github.com/SAA-SDT/TS-EAS-subteam-notes/wiki/Design-Principles). TS-EAS has a newly established set of design principles which feed into all of the work we do, and help to make things consistent across all of the work we do. So this includes things like controlling the sequence where child elements might appear within a parent element. In the EAC-CPF specific context, this has also defined a couple of plural elements such as <functions> and <places>, which group singular elements underneath them. So a singular function <placeDoc> element can be included within the plural element. The second example is for the <...Set> elements, which can be used to group elements with different concepts.

So in the example of <dateSet>, you can have <date> underneath that or multiple date elements, or <dateRange>, or a combination of the two.

We've been working on the EAC-CPF revision, but there's also been work on more closely aligning the EAD and EAC-CPF schemas as a whole. This would address the considerable overlap between the standards, make them easier to maintain, and easier to teach, learn and use. It's also going to look at making sure that anywhere we have elements with the same names across both schemes, that they are defined in the same way, and used in the same way. This has resulted in the renaming of some elements and attributes.

next slide

- Renamed as part of alignment
    - <abbreviation> to <shortCode>
    - <citation> to <reference>
    - <placeEntry> to <placeName>
    - <script> to <writingSystem>

I probably should emphasize that the changes will only be applied to EAC-CPF in this revision, and subsequently

when EAD has a revision then they will move their form into that one, so EAC-CPF will come first.

- Renamed to be more precise
    - <eac-cpf> to <eac>
    - <nameEntryParallel> to <nameEntrySet>
    - <entityId> to <identityId>

next slide

<reference> has replaced the <sourceEntry> element. The <outline> and <level> elements, as well as the <objectBinWrap> element have been removed from the schema. In order to represent data in hierarchies of more complex lists, the <list> element can now be nested within other lists.

next slide

This slide shows an example of a series of nested lists used to encode a genealogy. It also shows the new <head> element that can be used to encode a title or caption.

next slide

Following the lead made by EAD in EAD3, XML namespaces and XLInk namespaces have either been replaced, for example the @xml:id attribute has been replaced by an @id attribute, and some Xlink attributes have been removed.

next slide

There are some optional attributes that have been introduced for all elements—@id and @target can be used for internal linking, and there will be more information about using them shortly, and @audience can be used to specify where content should be visible to all viewers, or only available to repository staff.

The optional @languageOfElement and @scriptOfElement attributes have been introduced to all elements that have content to specify the language and/or script of the contents of the element. These can be particularly useful in multi-lingual EAC-CPF instances, or where instances include translations of some content.

next slide

[Speaker: Silke Jagodzinski] The control area is mainly updated for EAD alignment and to simplify the schema.

next slide

Changes due to EAD alignment are new optional attributes @countryEncoding, @dateEncoding, @languageEncoding, @repositoryEncoding, and @scriptEncoding in the element <control> for external encoding standards. And we also added the new optional element <representation> as a child element of <control>.

next slide

Furthermore, we have some elements that were transformed to attributes as we realised they are more property information about the EAC instance than data, which means they are more likely data about data.

See our solution: there are elements to encode status information on an EAC-CPF instance, like <maintanceStatus> and <publicationStatus>. Both become attributes in the <control> element as they are properties of this EAC instance but not data about this EAC file.

next slide

Example slide with the new attributes, more attributes are possible:

```
<control audience="external" base="baseURI" countryEncoding="iso3166-1" dateEncoding="iso8601"
id="control1" languageEncoding="iso639-1" maintenanceStatus="new" publicationStatus="inProcess"
repositoryEncoding="iso15511" scriptEncoding="iso15924">
```

next slide

There are also two elements with type information that were turned into attributes. Within the maintenance event to describe the maintenance history of the EAC instance, the elements <eventType> and <agentType> become attributes of the elements <maintenanceEvent> and <agent>.

next slide

Example slide with the new attributes for maintenance event:

```
<maintenanceEvent maintenanceEventType="created">
    <agent agentType="human">TS EAS EAC-CPF team</agent>
    <eventDateTime standardDateTime="2021-04-027T08:01:48"> 27.04.2021</eventDateTime>
    <eventDescription>example for webinar</eventDescription>
```

</maintenanceEvent>

With that solution, four elements in control vanished and were transformed into attributes.

next slide

Within the control area we also emphasized existing elements. The convention declaration in control will be used to provide conventions and rules for transliteration and for any forms of names.

Information given in the elements <source> and <maintenanceEvent> lay the foundation for the description of an assertion to the EAC-CPF description. Ailie will explain that later on.

The optional attribute @localType is used to support an element semantically. The specification of values available in @localType is given in the <localTypeDeclaration>.

With some new optional reference attributes we're able to link to a specific convention declaration, maintenance event, source, and local type declaration. In detail: we are able to link to the @id attributes of the respective elements.

The attributes names become one of our favorites in the new schema and you'll see some use cases later on.

Ailie will say more on linking and referencing now.

next slide

next slide

[Speaker: Ailie Smith] The @valueURI, @vocabularySource and @vocabularySourceURI attributes are available on a number of elements for references to things like external vocabularies or ontologies.

To reference external sources for the contents of the EAC-CPF instance, use the <reference> element within <source> in the <control> area of the instance. <reference> can include an @href attribute to point to a specific address for a remote resource. The <reference> element is also available within other elements, such as <event> or <abstract> to reference external resources for context.

next slide

In this example here, we've got a few different ways to do the external referencing. So the top case, it's using the <control> area to link to an external source, which is the Barack Obama Presidential Library, and that includes the

@href attribute there. In the second example, this is a name entry for Barack Obama, it's using the @vocabularySource URI to point to Wikidata and the @valueURI to point to the specific value where this name entry has come from. And in the third example, it's using a <relation> element where there is a <reference> at the bottom that's pointing to an external resource.

next slide

There are some enhanced options for creating internal references within a single EAC-CPF instance. The @id attribute is available on all elements to assign an identifier that is unique within the instance. There is also a new @target attribute that can be used to refer to the @id of another element. You can also reference specific elements in <control> from descriptive elements in an EAC-CPF instance using specific attributes, I'll be talking about this in the next section, but first we can have a look at an example of creating internal references using the @id and @target attributes.

next slide

So this is a snippet of data not a complete record, but at the top you've got an <occupation>, where the place is found, and that place has a @target attribute of #place1. So what that's doing is linking that place name there to the @id attribute that appears below on the <place> record there, where Bern is described in a lot more detail. So you can link between the two sections in a single EAC-CPF instance.

next slide

Based on user feedback, we have worked on including evidence-based assertions in this version of EAC-CPF. So for statements that form part of the EAC-CPF description, you can encode who added the assertion, based on which source, and following which rules. This is especially useful where they might be conflicting statements in an EAC-CPF description, such as different spellings of names, or different dates of births, that have come from different sources. The new @maintenanceEventReference, @sourceReference and @conventionDeclaration references allow you to refer from a descriptive element in the EAC-CPF instance back to <maintenanceEvent>, <source> or <conventionDeclaration> statements in the control section of the EAC-CPF instance, and you can refer to more than one of these.

next slide

So the best thing to do is look at an example of this. So in this case we have a <maintenanceEvent>, which would be

up in the <control> section, where John Smith has made an update on the 23rd of February, this year. We have a <source>, which is pointing to the History of the University of Oregon. We also have a specific page in that reference, where this information comes from. The bottom is the <relation>, which would be further down in the record. And this <relation> is pointing to the source of this information and who updated the record to include this. So that's all included within the <relation>, and you can trace that which is very useful.

next slide

[Speaker: Silke Jagodzinski] Encoding various forms of names is essential for EAC-CPF producers. There are different reasons that make it necessary to encode several names for one entity.

next slide

Feedback from the user community showed that the name encoding was too elaborate for fast and simple encoding.

First we renamed the element <nameEntryParallel> to <nameEntrySet>. It can still contain one or more name entries and dates. You can still also use a single name entry for an entity without this wrapper element if needed.

Similar to our proposal in the control area, we transformed some elements to attributes, as we think they are properties of a name. The elements <authorizedForm> and <alternativeForm>, used to identify the status of the name according to a particular set of rules, were transferred to an optional @status attribute with the values authorized or alternative in the <nameEntry> element. The element <preferredForm> was turned into an attribute with the same name with Boolean expression true or false.

Rules or conventions to express the name or according to which the name is the authorized or alternative form can be defined in <conventionDeclaration> above in the control area; by the way the same applies to any rule or convention. The new optional attribute @conventionDeclarationReference in <nameEntry> points directly to the according convention declaration (i.e., to <conventionDeclaration @id>).

next slide

Example slide: it is a snippet, please imagine more information around

```
<conventionDeclaration id="cd2">
        <reference>[...] AFNOR NF Z44-060 [...] </reference>
</conventionDeclaration>
```

[...]

```
<nameEntry preferredForm="true" status="authorized" conventionDeclarationReference="#cd2">
        <part>Institut international des droits de l'homme</part>

</nameEntry>
```

next slide

Example slide: again it is a snippet, imagine more translations of the name and use dates within this name entry set.

Use the well known attribute @localType to indicate if various names are parallel names, translations, or if they are native names.

```
<nameEntrySet localType="parallel">
        <nameEntry languageOfElement="de" preferredForm="true" status="authorized" localType="native">
                <part localType="surname">Arendt</part>
                <part localType="firstname">Hannah</part>
        </nameEntry>
        <nameEntry languageOfElement="ja" scriptOfElement="Jpan" preferredForm="false"
        status="authorized" localType="translation">
                <part localType="surname">アーレント</part>
                <part localType="firstname">ハナ</part>
        </nameEntry> [...]
</nameEntrySet>
```

next slide

And now it's Ailie again for new place encoding.

next slide

[Speaker: Ailie Smith] The places can be fully encoded with the <place> element, within the plural <places> element, as well as within <relation> and within structured chronological list of events, within the <chronItem> or <chronItemSet> elements.

<placeEntry> has been renamed <placeName>. Individual <placename>s can be included within singular elements such as <function>, <occupation> and <mandate> without a <place> wrapper element and associated encoding.

next slide

The <place> element now requires at least one of name of place; role of place; a physical address; contact information; or geographic coordinates. As mentioned earlier, <placeEntry> has been renamed <placename> and this element is highly recommended to be included within <place>. An addition to this version of EAC-CPF is the new <contact> element. This works much like the existing <address> element, but where <address> encodes a physical address, <contact> is to be used for providing contact details, such as email addresses, phone numbers and web pages. Geographic coordinates can now be encoded in an element, rather than as @latitude, @longitude and @altitude attributes. Date information and additional descriptive information can also be included in <place>.

next slide

And here we have another snippet-type example. It's actually a fairly complete record for a place. So we have a <placeName>—the Tokyo Imperial Palace. It has a <placeRole>, includes <geographicCoordinates>, includes the street address for the Imperial Palace, and in this case it uses the new <contact> element to include the homepage for the Tokyo Imperial Palace.

next slide

There have been some changes in the way that dates can be encoded. This is based both on alignment with EAD, and on feedback from EAC-CPF users on needing the ability to encode uncertain and unknown dates.

next slide

To achieve this EAC-CPF is adopting the @certainty attribute from EAD as well as @calendar and @era to add additional information about dates. EAC-CPF will also enable the use of the Extended Date/Time format, which has been included in the latest version of the ISO 8601 standard. A new @status attribute has also been added to <fromDate> and <toDate> allow for indicating where parts of a date range may be unknown, or where a date range is ongoing.

next slide

Some examples here, we've got snippets. So in the first one, using the @certainty attribute to say that it's a circa date. We know its around 1789 but we don't know for certain, and that lets you encode that. In the second example, it's a <dateRange>. And this is a <dateRange> where the <fromDate> or the start of the <dateRange> is unknown. So it's using a @status value of "unknown" to specify that. And the <toDate>, or the end of the <dateRange> is uncertain, so in the text within the tags it's got circa 210, however, in the @standardDate it's using

0210?, and that question mark is the EDTF value that indicates uncertainty there. The third example is a <dateSet>, which includes a single date with a @standardDate on it, and a <dateRange> as well where the end of the <dateRange>, or the <toDate>, is ongoing. So the date range, it started but it might be an organization that is still continuing to operate, so it's saying that it is an ongoing <dateRange> and there is actually no <toDate>.

next slide

[Speaker: Silke Jagodzinski] Relations got a new encoding structure to simplify and for better interoperability. The relation encoding is more generic now.

next slide

The optional element <relations> still serves as a wrapper element for one or more single relation entries. But we removed the elements <functionRelation>, <resourceRelation>, and <cpfRelation> in favor of a generic singular element <relation>.

Each element <relation> contains at least one required element <targetEntity> to identify the related entity. An attribute @targetType specifies the type of the entity as an agent, meaning a corporate body, a person, or a family. The target type could also be a resource or a function. The targeted entity can be named with text in the element content and/or encoded with URIs and references to an authority record or any other external file using the new optional attributes @vocabularySource, @vocabularyURI, and @vocabularySourceURI which Ailie mentioned before.

The optional element <relationType> within <relation> specifies the type of relation that the EAC-CPF entity has to the targeted entity. Relation types are not fixed and can be given as identity, hierarchical, temporal, family, or associative, but also as creator or subject of a resource. Relations between the entity being described and a function could be controls, owns, or performs.

A third new target element is called <targetRole> and can be used to provide information about the role of the targeted entity toward the entity described. Family roles like parent, child, or sibling can be given in that text element.

Furthermore, relation context can be given with date, place, and descriptive information in the according child elements.

next slide

Example slide:

```
<relations>
        <relation id="r1">
            <targetEntity vocabularySource="GND" vocabularySourceURI="http://d-nb.info/gnd/"
              valueURI="http://d-nb.info/gnd/1190671301" targetType="person">
                <part localType="family">Arendt</part>
                <part localType="given">Paul</part>
            </targetEntity>
            <relationType languageOfElement="en">Family</relationType>
            <targetRole languageOfElement="en">parent</targetRole>
            </relation>
    </relations>
```

next slide

As previously mentioned, more and detailed information, some example encodings, and all files are available on our standards homepage eac.staatsbibliothek-berlin.de menu item Schema Revision 2021.

next slide

Again I'd like to ask you to take a look at our questionnaire. We're especially interested in your opinion on language encoding and usage of namespaces within your XML files. In terms of schema alignment, mentioned by Ailie before, the results may affect the upcoming EAD revision.

If you are already using EAC-CPF we would like to see your files for encoding examples. We would really like to see how somebody is using multiple identities. We would also like to see evidence-based assertions. So please take a look at our questionnaire to learn about the areas of EAC-CPF we've identified as fuzzy.

next slide

Enough talk. It is your turn to ask any question or to give us your opinion on this present EAC-CPF 2.0 draft in our Q&A session.

[Speaker: Cory Nimer] Thank you so much, Silke and Ailie. So we'll go ahead and move into our Q&A session and stop the recording at this point.